math612.4 612.4 Linear algebra, multivariate calculus and multivariate statistics

Gunnar Stefansson

19. desember 2016

Acknowledgements
MareFrame is a EC-funded RTD project which seeks to remove the barriers preventing more widespread use of the ecosystem-based approach to fisheries management.
`http://mareframe-fp7.org`
This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no.613571.
`http://mareframe-fp7.org`
Háskóli Íslands
`http://www.hi.is/`

Efnisyfirlit

1 Vectors and Matrix Operations

1.1 Numbers, vectors, matrices

Recall that the set of real numbers is $\mathbb{R}$ and that a vector , $v \in \mathbb{R}^n$ is just an n-tuple of numbers.

Similarly, an *nxm* matrix is just a table of numbers, with n rows and m columns and we can write

$$A_{mn} \in \mathbb{R}^{mn}$$

Note that a vector is normally considered equivalent to a $n \times 1$ matrix i.e. we view these as column vectors.

1.1.1 Examples

Example 1.1. In R, a vector can be generated with:

```
X<- 3:6
X
[1] 3 4 5 6
```

A matrix can be generated in R as follows,

```
matrix(X)
   [,1]
[1,] 3
[2,] 4
[3,] 5
[4,] 6
```

Note 1.1. We note that R distinguishes between vector and matrices.

1.2 Elementary Operations

We can define multiplication of a real number k and a vector $v = (v_1, \ldots, v_n)$ by $k \cdot v = (kv_1, \ldots, kv_n)$. The sum of two vectors in $\mathbb{R}^n$, $v = (v_1, \ldots, v_n)$ and $u = (u_1, \ldots, u_n)$ as the vector $v + u = (v_1 + u_1, \ldots, v_n + u_n)$. We can define multiplication of a number and a matrix and the sum of two matrices (of the same sizes) similarly.

1.2.1 Examples

Example 1.2. `A <- matrix(c(1,2,3,4), nr=2, nc=2)`

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

```
B <- matrix(c(1,0,2,1), nr=2, nc=2)
B
     [,1] [,2]
[1,] 1 2
[2,] 0 1

A+B
  [,1] [,2]
[1,] 2 5
[2,] 2 5
```

1.3 The tranpose of a matrix

In R, matrices may be constructed using the "matrix"function and the transpose of A, A', may be obtained in R by using the "t"function:

```
A<-matrix(1:6, nrow=3)
t(A)
```

1.3.1 Details

If A is an $n \times m$ matrix with element a_{ij} in row i and column j, then A' or A^T is the $m \times n$ matrix with element a_{ij} in row j and column i.

1.3.2 Examples

Example 1.3. Consider a vector in R

```
x<-1:4
x
[1] 1 2 3 4
t(x)
     [,1] [,2] [,3] [,4]
[1,] 1 2 3 4
matrix(x)
     [,1]
[1,] 1
[2,] 2
[3,] 3
[4,] 4
t(matrix(x))
     [,1] [,2] [,3] [,4]
[1,] 1 2 3 4
```

Note 1.2. Note that the first solution gives a $1 \times n$ matrix and the second solution gives a $n \times 1$ matrix.

1.4 Matrix multiplication

Matrices A and B can be multiplied together if A is an $n \times p$ matrix and B is an $p \times m$ matrix. The general element $c_i j$ of $n \times m$; $C = AB$ is found by pairing the $i^t h$ row of C with the $j^t h$ column of B, and computing the sum of products of the paired terms.

$$\underbrace{\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}}_{3\times 2} \underbrace{\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}}_{2\times 2} = \begin{bmatrix} 1\cdot 1+2\cdot 1 & 1\cdot 2+2\cdot 3 \\ 3\cdot 1+4\cdot 1 & 3\cdot 2+4\cdot 3 \\ 5\cdot 1+6\cdot 1 & 5\cdot 2+6\cdot 3 \end{bmatrix} = \underbrace{\begin{bmatrix} 3 & 8 \\ 7 & 18 \\ 11 & 28 \end{bmatrix}}_{3\times 2}$$

1.4.1 Details

Matrices A and B can be multiplied together if A is a $n \times p$ matrix and B is a $p \times m$ matrix. Given the general element c_{ij} of nxm matrix, $C = AB$ is found by pairing the $i^t h$ row of C with the $j^t h$ column of B, and computing the sum of products of the paired terms.

1.4.2 Examples

Example 1.4. Matrices in R

```
A<-matrix(c(1,3,5,2,4,6),3,2)
A
     [,1] [,2]
[1,] 1 2
[2,] 3 4
[3,] 5 6
B<-matrix(1,1,2,3)2,2)

B<-matrix(c(1,1,2,3),2,2)
 B
     [,1] [,2]
[1,] 1 2
[2,] 1 3
A%*%B
     [,1] [,2]
[1,] 3 8
[2,] 7 18
[3,] 11 28
```

1.5 More on matrix multiplication

Let A, B, and C be $m \times n$, $n \times l$, and $l \times p$ matrices, respectively. Then we have

$$(AB)C = A(BC).$$

In general, matrix multiplication is not commutative, that is $AB \neq BA$.
We also have

$$(AB)' = B'A'.$$

In particular, $(Av)'(Av) = v'A'Av$, when v is a $n \times 1$ column vector.

More obvious are the rules

1. $A + (B + C) = (A + B) + C$
2. k(A+B)=kA+kB
3. A(B+C)=AB+AC,

where $k \in \mathbb{R}$ and when the dimensions of the matrices fit.

1.6 Linear equations

1.6.1 Details

Detail:
General linear equations can be written in the form $Ax = b$.

1.6.2 Examples

Example 1.5. The set of equations

$2x + 3y = 4$
$3x + y = 2$

can be written in matrix formulation as

$$\begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

i.e. $A\underline{x} = \underline{b}$ for an appropriate choice of of $A, \underline{x}$ and $\underline{b}$

1.7 The unit matrix

The n x n matrix

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}$$

is the identity matrix. This is because if a matrix $\mathbf{A}$ is n x n then $\mathbf{AI} = \mathbf{A}$ and $\mathbf{IA} = \mathbf{A}$

1.8 The inverse of a matrix

If A is an $n \times n$ matrix and B is a matrix such that

$$BA = AB = I$$

Then B is said to be the inverse of A, written

$$B = A^{-1}$$

Note that if A is an $n \times n$ matrix for which an inverse exists, then the equation $Ax = b$ can be solved and the solution is $x = A^{-1}b$.

1.8.1 Examples

Example 1.6. If matrix A is:
$\begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix}$
then A^{-1} is:
$\begin{bmatrix} \frac{-1}{4} & \frac{3}{4} \\ \frac{3}{4} & \frac{1}{2} \end{bmatrix}$

2 Some notes on matrices and linear operators

2.1 The matrix as a linear operator

Let A be an $m \times n$ matrix, the function

$$T_A : \mathbb{R}^n \rightarrow \mathbb{R}^m, T_A(\underline{x}) = A\underline{x},$$

is linear, that is

$$T_A(a\underline{x} + b\underline{y}) = aT_A(\underline{x}) + bT_A(\underline{y})$$

if $\underline{x}, \underline{y} \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$.

2.1.1 Examples

Example 2.1. If A= $\begin{bmatrix}1 & 2\end{bmatrix}$ then $T_A(\underline{x}) = x + 2y$ where $\underline{x} = \binom{x}{y} \in \mathbb{R}^2$

Example 2.2. If A= $\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$ then $T_A\binom{x}{y} = \begin{bmatrix}y\\x\end{bmatrix}$

Example 2.3. If A= $\begin{bmatrix}0 & 2 & 3\\1 & 0 & 1\end{bmatrix}$ then $T_A\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{bmatrix}2y+3z\\x+z\end{bmatrix}$

Example 2.4. If $T\binom{x}{y} = \begin{pmatrix}x+y\\2x-3y\end{pmatrix}$ then $T(\underline{x}) = A\underline{x}$ if we set A = $\begin{bmatrix}1 & 1\\2 & -3\end{bmatrix}$

2.2 Inner products and norms

Assuming x and y are vectors, then we define their inner product by

$$x \cdot y = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

where $x = \begin{pmatrix}x_1\\\vdots\\x_n\end{pmatrix}$ and $y = \begin{pmatrix}y_1\\\vdots\\y_n\end{pmatrix}$

2.2.1 Details

If $x, y \in \mathbb{R}^n$ are arbitrary (column) vectors, then we define their inner product by

$$x \cdot y = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

where $x = \begin{pmatrix}x_1\\\vdots\\x_n\end{pmatrix}$ and $y = \begin{pmatrix}y_1\\\vdots\\y_n\end{pmatrix}$.

Note 2.1. Note that we can also view x and y as $n \times 1$ matrices and we see that $x \cdot y = x'y$.

Definition 2.1. The normal length of a vector is defined by $\|x\|^2 = x \cdot x$. It may also be expressed as $\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$.

It is easy to see that for vectors a, b and c we have $(a+b) \cdot c = a \cdot c + b \cdot c$ and $a \cdot b = b \cdot a$.

2.2.2 Examples

Two vectors x and y are said to be orthogonal if $x \cdot y = 0$

Example 2.5. If $x = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ and $y = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, then

$$x \cdot y = 3 \cdot 2 + 4 \cdot 1 = 10,$$

and

$$\|x\|^2 = 3^2 + 4^2 = 25,$$

so

$$\|x\| = 5$$

2.3 Orthogonal vectors

Two vectors x and y are said to be orthogonal if $x \cdot y = 0$ denoted $x \perp y$

2.3.1 Details

Definition 2.2. Two vectors x and y are said to be **orthogonal** if $x \cdot y = 0$ denoted $x \perp y$

If $a, b \in \mathbb{R}^n$ then

$$\|a + b\|^2 = a \cdot a + 2a \cdot b + b \cdot b$$

so

$$\|a + b\|^2 = \|a\|^2 + \|b\|^2 + 2\underline{ab}.$$

Note 2.2. Note that if $a \perp b$ then $\|a + b\|^2 = \|a\|^2 + \|b\|^2$, which is Pythagoras' theorem in n dimensions.

2.4 Linear combinations of i.i.d. random variables

Suppose $X_1, \ldots., X_n$ are i.i.d. random variables and have mean $\mu_1, \ldots., \mu_n$ and variance σ^2 then the expected value of Y of the linear combination is

$$Y = \sum a_i X_i$$

and if $a_1, \ldots., a_n$ are real constants then the mean is:

$$\mu_Y = \sum a_i \mu_i$$

and the variance is:

$$\sigma^2 = \sum a_i^2 \sigma_i^2$$

2.4.1 Examples

Example 2.6. Consider two i.i.d. random variables, Y_1, Y_2 and a specific linear combination of the two, $W = Y_1 + 3Y_2$.

We first obtain

$$E[W] = E[Y_1 + 3Y_2] = E[Y_1] + 3E[Y_2] = 2 + 3 \cdot 2 = 2 + 6 = 8.$$

Similarly, we can first use independence to obtain

$$V[W] = V[Y_1 + 3Y_2] = V[Y_1] + V[3Y_2]$$

and then (recall that $V[aY] = a^2 V[Y]$)

$$V[Y_1] + V[3Y_2] = V[Y_1] + 3^2 V[Y_2] = 1^2 + 3^2 = 1(4) + 9(4) = 40$$

Normally, we just write this up in a simple sequence

$$V[W] = V[Y_1 + 3Y_2] = V[Y_1] + 3^2 V[Y_2] = 1^2 + 3^2 = 1(4) + 9(4) = 40$$

2.5 Covariance between linear combinations of i.i.d random variables

Suppose $Y_1, \ldots, Y_n$ are i.i.d., each with mean μ and variance σ^2 and $a, b \in \mathbb{R}^n$. Writing $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$, consider the linear combination $a'Y$ and $b'Y$.

2.5.1 Details

The covarience between random variables U and W is defined by

$$Cov(U, W) = E[(U - \mu_u)(W - \mu_w)]$$

where

$$\mu_u = E[U], \mu_w = E[W]$$

Now, let $U = a'Y = \sum Y_i a_i$ and $W = b'Y = \sum Y_i b_i$, where $Y_1, \ldots, Y_n$ are i.i.d. with mean μ and variance σ^2, then we get

$$\begin{aligned} Cov(U, W) &= E[(\underline{a}'Y - \Sigma a_\mu)(\underline{b}'Y - \Sigma b\mu)] \\ &= E[(\Sigma a_i Y_i - \Sigma a_i \mu)(\Sigma b_j Y_j - \Sigma b_j \mu)] \end{aligned}$$

and after some tedious (but basic) calculations we obtain

$$Cov(U, W) = \sigma^2 a \cdot b$$

2.5.2 Examples

Example 2.7. If Y_1 and Y_2 are i.i.d., then

$$Cov(Y_1+Y_2, Y_1-Y_2) = Cov((1,1)\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, (1,-1)\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix})$$
$$= (1,1)\begin{pmatrix} 1 \\ -1 \end{pmatrix}\sigma^2$$
$$= 0$$

and in general, $Cov(\underline{a}'\underline{Y}, \underline{b}'\underline{Y}) = 0$ if $\underline{a} \perp \underline{b}$ and $Y_1, \ldots, Y_n$ are independent.

2.6 Random vectors

$Y = (Y_1, \ldots, Y_n)$ is a random vector if $Y_1, \ldots, Y_n$ are random variables.

2.6.1 Details

Definition 2.3. If $EY_i = \mu_i$ then we typically write

$E(Y) = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \mu$

If $Cov(Y_i, Y_j) = \sigma ij$ and $V[Y_i] = \sigma_{ii} = \sigma_i^2$, then we define the matrix

$$\Sigma = (\sigma_{ij})$$

containing the variances and covariances. We call this matrix the **covariance matrix** of Y, typically denoted $V[Y] = \Sigma$ or $Cov[Y] = \Sigma$.

2.6.2 Examples

Example 2.8. If $Y_i, \ldots, Y_n$ are i.i.d., $EY_i = \mu$, $VY_i = \sigma^2$, $a, b \in \mathbb{R}^n$ and $U = a'Y$, $W = b'Y$, and T= $\begin{bmatrix} U \\ W \end{bmatrix}$

then

ET= $\begin{bmatrix} \Sigma a_i \mu \\ \Sigma b_i \mu \end{bmatrix}$

VT= Σ = $\sigma^2 \begin{bmatrix} \Sigma a_i^2 & \Sigma a_i b_i \\ \Sigma a_i b_i & \Sigma b_i^2 \end{bmatrix}$

Example 2.9. If $\underline{Y}$ is a random vector with mean μ and variance-covariance matrix Σ, then

$$E[a'Y] = a'\mu$$

and

$$V[a'Y] = a'\Sigma a.$$

2.7 Transforming random vectors

Suppose

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

is a random vector with $E\mathbf{Y} = \mu$ and $V\mathbf{Y} = \Sigma$ where the variance-covariance matrix

$$\Sigma = \sigma^2 \mathbf{I}$$

2.7.1 Details

Note that if $Y_1, \ldots, Y_n$ are independent with common variance σ^2 then

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \ldots & \sigma_{2n} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \ldots & \sigma_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \sigma_{n3} & \ldots & \sigma_n^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1^2 & 0 & \ldots & \ldots & 0 \\ 0 & \sigma_2^2 & \ddots & 0 & \vdots \\ \vdots & \ddots & \sigma_3^2 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & 0 & \sigma_n^2 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} 1 & 0 & \ldots & \ldots & 0 \\ 0 & 1 & \ddots & 0 & \vdots \\ \vdots & \ddots & 1 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & 0 & 1 \end{bmatrix} = \sigma^2 \mathbf{I}$$

If A is an m x n matrix, then

$$E[A\mathbf{Y}] = A\mu$$

and

$$V[A\mathbf{Y}] = A\Sigma A'$$

3 Ranks and determinants

3.1 The rank of a matrix

The rank of an *nxp* matrix, A, is the largest number of columns of A, which are not linearly dependent (i.e. the number of linearly independent columns).

3.1.1 Details

Vectors $a_1, a_2, \ldots, a_n$ are said to be linearly dependent if the constant $k_1, \ldots, k_n$ exists and are not all zero, such that

$$k_1\mathbf{a}_1 + k_2\mathbf{a}_2 + \ldots + k_n\mathbf{a}_n = 0$$

Note that if such constants exist, then we can write one of the a's as a linear combination of the rest, e.g. if $k_1 \neq 0$ then

$$a_1 = \mathbf{c_1} = -\frac{k_2}{k_1}a_2 - \ldots - \frac{k_2}{k_1}a_n$$

It can be shown that the rank of A is the same as the rank of A' i.e. the maximum number of linearly independent rows of A.

Note 3.1. Note that if rank $(A) = p$, then the columns are linearly independent.

3.1.2 Examples

Example 3.1. If

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

the rank of A = 2, since

$$k_1\begin{pmatrix} 1 \\ 0 \end{pmatrix} + k_2\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

if and only if

$$\begin{pmatrix} k_1 \\ k_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

so the columns are linearly independent.

Example 3.2. If

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

the rank of A = 2.

Example 3.3. If

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

the rank of A = 2, since

$$1\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 0\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + (-1)\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 0$$

(and hence the rank can not be more than 2) but

$$k_1\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + k_2\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

if and only if $k_1 = k_2 = 0$ (and hence the rank must be at least 2).

3.2 The determinant

Recall that for a 2x2 matrix,
$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$
the inverse of A is
$A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix}$

3.2.1 Details

Definition 3.1. The number $ad - bc$ is called the **determinant**of the 2x2 matrix A.

Definition 3.2. Now suppose A is an nxn matrix. An **elementary product** from the matrix is a product of n terms based on taking exactly one term from each column of row x. Each such term can be written in the form $a_{1j_1} \cdot a_{2j_2} \cdot a_{3j_3} \cdot \ldots \cdot a_{nj_n}$ where $j_1, \ldots, j_n$ is a permutation of the integers $1, 2, \ldots, n$. Each permutation σ of the integers $1, 2, \ldots, n$ can be performed by repeatedly interchanging two numbers.

Definition 3.3. A **signed elementary product** is an elementary product with a positive sign if the number of interchanges in the permutation is even but negative otherwise.

The determinant of A, det(A) or $|A|$ is the sum of all signed elementary products.

3.2.2 Examples

Example 3.4. $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$
then
$|A| = a_{1\underline{1}}a_{2\underline{2}} - a_{1\underline{2}}a_{2\underline{1}}$.

Example 3.5. $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$
$|A|$
= $a_{11}a_{22}a_{33}$ This is the identity permutation and has positive sign
$-a_{11}a_{23}a_{32}$ This is the permutation that only interchanges 2 and 3
$-a_{12}a_{21}a_{33}$ Only one interchange
$+a_{12}a_{23}a_{31}$ Two interchanges
$+a_{13}a_{21}a_{32}$ Two interchanges
$-a_{13}a_{22}a_{31}$ Three interchanges

Example 3.6. $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$
$|A| = -1$

Example 3.7. $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$
$|A| = 1 \cdot 2 \cdot 3 = 6$

Example 3.8. $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 3 & 0 \end{bmatrix}$
$|A| = 0$

Example 3.9. $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 3 & 0 \end{bmatrix}$
$|A| = -6$

Example 3.10. $A = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}$

$|A| = 0$

Example 3.11. $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

$|A| = 0$

3.3 Ranks, inverses and determinants

The following statements are true for an $n \times n$ matrix A:

- $rank(A) = n$
- $det(A) \neq 0$
- A has an inverse

3.3.1 Details

Suppose A is an $n \times n$ matrix. Then the following are truths:

- $rank(A) = n$
- $det(A) \neq 0$
- A has an inverse

4 Multivariate calculus

4.1 Vector functions of several variables

A vector-valued function of several variables is a function

$$f : \mathbb{R}^m \to \mathbb{R}^n$$

i.e. a function of m dimensional vectors, which returns n dimensional vectors.

4.1.1 Examples

Example 4.1. A real valued function of many variables: $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x_1, x_2, x_3) = 2x_1 + 3x_2 + 4x_3$.

Note 4.1. Note that f is linear and $f(x) = Ax$ where $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ and $A = \begin{bmatrix} 2 & 3 & 4 \end{bmatrix}$.

Example 4.2. Let

$$f : \mathbb{R}^2 \to \mathbb{R}^2$$

where:

$$f(x_1, x_2) = \begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \end{pmatrix}$$

Note 4.2. Note that $f(x) = Ax$, where $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

Example 4.3. Let

$$f : \mathbb{R}^3 \to \mathbb{R}^4$$

be defined by

$$f(x) = \begin{pmatrix} x_1 + x_2 \\ x_1 - x_3 \\ y - z \\ x_1 + x_2 + x_3 \end{pmatrix}$$

Note 4.3. Note that:

$$f(x) = Ax$$

where

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

Example 4.4. These multi-dimensional functions do not have to be linear, for example the function $f : \mathbb{R}^2 \to \mathbb{R}^2$

$$f(x) = \begin{pmatrix} x_1 x_2 \\ x_1^2 + x_2^2 \end{pmatrix},$$

is obviously not linear.

4.2 The gradient

Suppose the real valued function $f : \mathbb{R}^m \to \mathbb{R}$ is differentiable in each coordinate. Then the gradient of f, denoted ∇f is given by

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \quad \ldots \quad , \frac{\partial f}{\partial x_1} \right).$$

4.2.1 Details

Definition 4.1. Suppose the real valued function $f : \mathbb{R}^m \to \mathbb{R}$ is differentiable in each coordinate. Then the **gradient** of f, denoted ∇f is given by

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \quad \ldots \quad, \frac{\partial f}{\partial x_1}\right),$$

where each partial derivative $\frac{\partial f}{\partial x_i}$ is computed by differentiating f with respect to that variable, regarding the others as fixed.

4.2.2 Examples

Example 4.5.

$$f(\underline{x}) = x^2 + y^2 + 2xy;\ \frac{\partial f}{\partial x} = 2x + 2y, \frac{\partial f}{\partial y} = 2y + 2x, \nabla f = \begin{pmatrix} 2x + 2y, & 2y + 2x \end{pmatrix}$$

Example 4.6.

$$f(\underline{x}) = x_1 - x_2; \nabla f = \begin{pmatrix} 1, & -1 \end{pmatrix}$$

4.3 The Jacobian

Now consider a function $f : \mathbb{R}^m \to \mathbb{R}^n$. Write f_i for the i^{th} coordinate of f, so we can write $f(x) = (f_1(x), f_2(x), \ldots, f_n(x))$, where $x \in \mathbb{R}^m$. If each coordinate function f_i is differentiable in each variable we can form the *Jacobian matrix* of f:

$$\begin{pmatrix} \nabla f_1 \\ \vdots \\ \nabla f_n \end{pmatrix}.$$

4.3.1 Details

Now consider a function $f : \mathbb{R}^m \to \mathbb{R}^n$. Write f_i for the i^{th} coordinate of f, so we can write $f(x) = (f_1(x), f_2(x), \ldots, f_n(x))$, where $x \in \mathbb{R}^m$. If each coordinate function f_i is differentiable in each variable we can form the *Jacobian matrix* of f:

$$\begin{pmatrix} \nabla f_1 \\ \vdots \\ \nabla f_n \end{pmatrix}.$$

In this matrix, the element in the i^th row and j^th column is $\frac{\partial f_i}{\partial x_j}$.

4.3.2 Examples

Example 4.7. For the function

$$f(x,y) = \begin{pmatrix} x^2+y \\ xy \\ x \end{pmatrix} = \begin{pmatrix} f_1(x,y) \\ f_2(x,y) \\ f_3(x,y) \end{pmatrix},$$

the Jacobian matrix of f is the matrix

$$J = \begin{bmatrix} \nabla f_1 \\ \nabla f_2 \\ \nabla f_3 \end{bmatrix} = \begin{bmatrix} 2x & 2y \\ y & x \\ 1 & 0 \end{bmatrix}.$$

4.4 Univariate integration by substitution

If f is a continuous function and g is strictly increasing and differentiable then,

$$\int_{g(a)}^{g(b)} f(x)dx = \int_a^b f(g(t))g'(t)dt$$

4.4.1 Details

If f is a continuous function and g is strictly increasing and differentiable then,

$$\int_{g(a)}^{g(b)} f(x)dx = \int_a^b f(g(t))g'(t)dt$$

It follows that if X is a continuous random variable with density f and $Y = h(X)$ is a function of X that has the inverse $g = h^{-1}$, so $X = g(Y)$, then the density of Y is given by,

$$f_Y(y) = f(g(y))g'(y)$$

This is a consequence of

$$P[Y \leq b] = P[g(Y) \leq g(b)] = P[X \leq g(b)] = \int_{-\infty}^{g(b)} f(x)dx = \int_{-\infty}^{b} f(g(y))g'(y)dy.$$

4.5 Multivariate integration by substitution

Suppose f is a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^n$ is a one-to-one function with continuous partial derivatives. Then if $U \subseteq \mathbb{R}^n$ is a subset,

$$\int_{g(u)} f(\underline{x})d\underline{x} = \int_u (\underline{g}(\underline{y}))|J|d\underline{y}$$

where J is the Jacobian matrix and |J| is the absolute value of it's determinant.

$$J = \left| \begin{bmatrix} \frac{\partial g_1}{\partial y_1} & \frac{\partial g_1}{\partial y_2} & \cdots & \frac{\partial g_1}{\partial y_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial g_n}{\partial y_1} & \frac{\partial g_n}{\partial y_2} & \cdots & \frac{\partial g_n}{\partial y_n} \end{bmatrix} \right| = \left| \begin{bmatrix} \nabla g_1 \\ \vdots \\ \nabla g_n \end{bmatrix} \right|$$

4.5.1 Details

Suppose f is a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^n$ is a one-to-one function with continuous partial derivatives. Then if $U \subseteq \mathbb{R}^n$ is a subset,

$$\int_{g(u)} f(\underline{x})d\underline{x} = \int_u (\underline{g}(\underline{y}))|J|d\underline{y}$$

where J is the Jacobian determinant and |J| is its absolute value.

$$J = |\begin{bmatrix} \frac{\partial g_1}{\partial y_1} & \frac{\partial g_1}{\partial y_2} & \cdots & \frac{\partial g_1}{\partial y_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial g_n}{\partial y_1} & \frac{\partial g_n}{\partial y_2} & \cdots & \frac{\partial g_n}{\partial y_n} \end{bmatrix}| = |\begin{bmatrix} \nabla g_1 \\ \vdots \\ \nabla g_n \end{bmatrix}|$$

Similar calculations as in 4.5 give us that if X is a continuous multivariate random variable, $X = (X_1, \ldots, X_n)'$ with density f and $\underline{Y} = \underline{h}(\underline{X})$, where $\underline{\text{h}}$ is 1-1 with inverse $g = h^{-1}$. So, $\underline{X} = g(\underline{Y})$, then the density of $\underline{\text{Y}}$ is given by;

$$f_Y(\underline{y}) = f(g(y))|J|$$

4.5.2 Examples

Example 4.8. If $\underline{Y} = A\underline{X}$ where A is an $n \times n$ matrix with $det(A) \neq 0$ and $X = (X_1, \ldots, X_n)'$ are i.i.d. random variables, then we have the following results:

The joint density of $X_1 \cdots X_n$ is the product of the individual (marginal) densities,

$$f_X(\underline{x}) = f(x_1)f(x_2)\cdots f(x_n)$$

The matrix of partial derivatives corresponds to $\frac{\partial g}{\partial y}$ where $X = g(\underline{Y})$, i.e. these are the derivatives of the transformation: $\underline{X} = \underline{g}(\underline{Y}) = A^{-1}\underline{Y}$, or $\underline{X} = B\underline{Y}$ where $B = A^{-1}$.

But if $X = B\underline{Y}$, then

$$X_i = b_{i1}y_1 + b_{i2}y_2 + \cdots b_{ij}y_j \cdots b_{in}y_n$$

So, $\frac{\partial x_i}{\partial y_i} = b_{ij}$ and thus,

$$J = |\frac{\partial d\underline{x}}{\partial d\underline{y}}| = |B| = |A^{-1}| = \frac{1}{|A|}$$

The density of $\underline{\text{Y}}$ is therefore;

$$f_Y(\underline{y}) = f_X(g(y))|J| = f_X(A^{-1}\underline{y}) = |A^{-1}|$$

5 The multivariate normal distribution and related topics

5.1 Transformations of random variables

Recall that if X is a vector of continuous random variables with a joint probability density function and if $Y = h(X)$ such that h is a 1-1 function and continuously differentiable with inverse g so $X = g(Y)$, then the density of Y is given by

$$f_Y(y) = f(g(y))|J|$$

5.1.1 Details

J is the Jacobian determinant of g. In particular if $Y = AX$ then

$$f_Y(y) = f(A^{-1}y)|det(A^{-1})|$$

if A has an inverse.

5.2 The multivariate normal distribution

5.2.1 Details

Consider i.i.d. random variables, $Z_1,\ldots,Z_n \sim (0,1)$, written $\underline{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$ and let $\underline{Y} = A\underline{Z} + \underline{\mu}$ where A is an invertible nxn matrix and $\underline{\mu} \in \mathbb{R}^n$ is a vector, so $Z = A^{-1}(Y - \underline{\mu})$.

Then the p.d.f. of Y is given by

$$f_{\underline{Y}}(\underline{y}) = f_{\underline{Z}}(A^{-1}(\underline{y} - \underline{\mu}))|det(A^{-1})|$$

But the joint p.d.f. of $\underline{Z}$ is the product of the p.d.f.'s of $Z_1,\ldots,Z_n$, so $f_{\underline{Z}}(\underline{z}) = f(z_1)\cdot f(z_2)\cdot \ldots \cdot f(z_n)$ where

$$f(z_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}}$$

and hence

$$f_{\underline{Z}}(\underline{z}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{\frac{-z_i^2}{2}}$$

$$= (\frac{1}{\sqrt{2\pi}})^n e^{-\frac{1}{2}\sum_{i=1}^{n} z_i^2}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\underline{z}'\underline{z}}$$

since

$$\sum_{i=1}^{n} z_i^2 = \|\underline{z}\|^2 = \underline{z}\cdot\underline{z} = \underline{z}'\underline{z}$$

The joint p.d.f. of $\underline{Y}$ is therefore

$$f_{\underline{Y}}(\underline{y}) = f_{\underline{Z}}(A^{-1}(\underline{y} - \underline{\mu}))|det(A^{-1})|$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(A^{-1}(\underline{y}-\underline{\mu}))'(A^{-1}(\underline{y}-\underline{\mu}))} \frac{1}{|det(A)|}$$

We can write $det(AA') = det(A)^2$ so $|det(A)| = \sqrt{det(AA')}$ and if we write $\Sigma = AA'$, then

$$|det(A)| = |\Sigma|^{\frac{1}{2}}$$

Also, note that

$$(A^{-1}(\underline{y}-\underline{\mu}))'(A^{-1}(\underline{y}-\underline{\mu})) = (\underline{y}-\underline{\mu})'(A^{-1})'A^{-1}(\underline{y}-\underline{\mu}) = (\underline{y}-\underline{\mu})'\Sigma^{-1}(\underline{y}-\underline{\mu})$$

We can now write

$$f_{\underline{Y}}(\underline{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\underline{y}-\underline{\mu})\Sigma^{-1}(\underline{y}-\underline{\mu})}$$

This is the density of the multivariate normal distribution.
Note that

$$E[\underline{Y}] = \mu$$

$$V[\underline{Y}] = V[A\underline{Z}] = AV[\underline{Z}]A' = AIA' = \Sigma$$

Notation: $\underline{Y} \sim n(\underline{\mu}, \Sigma)$

5.3 Univariate normal transforms

The general univariate normal distribution with density

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

is a special case of the multivariate version.

5.3.1 Details

Further, if $Z \sim n(0,1)$, then clearly $X = aZ + \mu \sim n(\mu, \sigma^2)$ where $\sigma^2 = a^2$

5.4 Transforms to lower dimensions

If $Y \sim n(\mu, \Sigma)$ is a random vector of length n and A is an $m \times n$ matrix of rank $m \leq n$, then $AY \sim n(A\mu, A\Sigma A')$.

5.4.1 Details

If $Y \sim n(\mu, \Sigma)$ is a random vector of length n and A is an $m \times n$ matrix of rank $m \leq n$, then $AY \sim n(A\mu, A\Sigma A')$.

To prove this, set up an $(n-m) \times n$ matrix, B, so that the $n \times n$ matrix, C, formed from combining the rows of A and B is of full rank n. Then it is easy to derive the density of CY which also factors nicely into a product, only one of which contains AY, which gives the density for AY.

5.5 The OLS estimator

Suppose $Y \sim n(X\beta), \sigma^2 I)$. The ordinary least squares estimator, when the $n \times p$ matrix is of full rank, p, where $p \leq n$, is:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

The random variable which describes the process giving the data and estimate is:

$$b = (X'X)^{-1}X'Y$$

It follows that

$$\hat{\beta} \sim n(\beta, \sigma^2(X'X)^{-1})$$

5.5.1 Details

Suppose $Y \sim n(X\beta, \sigma^2 I)$. The ordinary least squares estimator, when the $n \times p$ matrix is of full rank, p, is:

$$\hat{\beta} = (X'X)^{-1}X'Y.$$

The equation below is the random variable which describes the process giving the data and estimate:

$$b = (X'X)^{-1}X'Y$$

If $B = (X'X)^{-1}X'$, then we know that

$$BY \sim n(BX\beta, B(\sigma^2 I)B')$$

Note that

$$BX\beta = (X'X)^{-1}X'X\beta = \beta$$

and

$$\begin{aligned} B(\sigma^2 I)B' &= \sigma(X'X)^{-1}X'[(X'X)^{-1}X']' \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

It follows that

$$\hat{\beta} \sim n(\beta, \sigma^2(X'X)^{-1})$$